

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Palese, Peter
O'Neill, Robert
- (ii) TITLE OF INVENTION: IDENTIFICATION AND USE OF ANTIVIRAL
COMPOUNDS THAT INHIBIT INTERACTION OF HOST CELL PROTEINS
AND VIRAL PROTEINS REQUIRED FOR VIRAL REPLICATION
- (iii) NUMBER OF SEQUENCES: 20
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Pennie & Edmonds
 - (B) STREET: 1155 Avenue of the Americas
 - (C) CITY: New York
 - (D) STATE: New York
 - (E) COUNTRY: USA
 - (F) ZIP: 10036-2711
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER: US 08/444,994
 - (B) FILING DATE: 19-MAY-1995
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Coruzzi, Laura A.
 - (B) REGISTRATION NUMBER: 30,742
 - (C) REFERENCE/DOCKET NUMBER: 6923-054
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: (212) 790-9090
 - (B) TELEFAX: (212) 869-9741/8864
 - (C) TELEX: 66141 PENNIE

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GCAAAGCAGG AGAAACCAC

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear
(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:
GGGTCCATCT GATAGATATG AGAG

24

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 48 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 36
(D) OTHER INFORMATION: /mod_base= i

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 37
(D) OTHER INFORMATION: /mod_base= i

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 41
(D) OTHER INFORMATION: /mod_base= i

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 42
(D) OTHER INFORMATION: /mod_base= i

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 46
(D) OTHER INFORMATION: /mod_base= i

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 47
(D) OTHER INFORMATION: /mod_base= i

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:
CUACUACUAC UAGGCCACGC GTCGACTACT ACGGGNNGG NNGGGNNG

48

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

20

TCCTGATGTT GCTGTAGACG

(2) INFORMATION FOR SEQ ID NO:5:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

20

GCACGACTAG TATGATTTGC

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 8 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Thr Gly Ala Gly Ala Gly Leu Gly
1 5

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 5 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Tyr Ser Ala Ala Lys
1 5

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 27 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 1..27

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

GAC TGG CTG GAA TTC CCC ATG GCG TCC
Asp Trp Leu Glu Phe Pro Met Ala Ser
1 5

27

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Asp Trp Leu Glu Phe Pro Met Ala Ser
1 5

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 2940 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: cDNA

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: 47..1663

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

CTAACTTCAG CGGTGGCACC GGGATCGGTT GCCTTGAGCC TGAAAT ATG ACC ACC 55
Met Thr Thr
1

CCA GGA AAA GAG AAC TTT CGC CTG AAA AGT TAC AAG AAC AAA TCT CTG 103
Pro Gly Lys Glu Asn Phe Arg Leu Lys Ser Tyr Lys Asn Lys Ser Leu
5 10 15

AAT CCC GAT GAG ATG CGC AGG AGG AGG GAG GAA GAA GGA CTG CAG TTA 151
Asn Pro Asp Glu Met Arg Arg Arg Arg Glu Glu Glu Gly Leu Gln Leu
20 25 30 35

CGA AAG CAG AAA AGA GAA GAG CAG TTA TTC AAG CGG AGA AAT GTT GCT 199
Arg Lys Gln Lys Arg Glu Glu Gln Leu Phe Lys Arg Arg Asn Val Ala
40 45 50

ACA GCA GAA GAA GAA ACA GAA GAA GAA GTT ATG TCA GAT GGA GGC TTT 247
Thr Ala Glu Glu Glu Thr Glu Glu Glu Val Met Ser Asp Gly Gly Phe
55 60 65

| | |
|---|------|
| CAT GAG GCT CAG ATT AGT AAC ATG GAG ATG GCA CCA GGT GGT GTC ATC His Glu Ala Gln Ile Ser Asn Met Glu Met Ala Pro Gly Gly Val Ile 70 75 80 | 295 |
| ACT TCT GAC ATG ATT GAG ATG ATA TTT TCC AAA AGC CCA GAG CAA CAG Thr Ser Asp Met Ile Glu Met Ile Phe Ser Lys Ser Pro Glu Gln Gln 85 90 95 | 343 |
| CTT TCA GCA ACA CAG AAA TTC AGC AAG CTG CTT TCA AAA GAA CCT AAC Leu Ser Ala Thr Gln Lys Phe Arg Lys Leu Leu Ser Lys Glu Pro Asn 100 105 110 115 | 391 |
| CCT CCT ATT GAT GAA GTT ATC AGC ACA CCA GGA GTA GTG GCC AGG TTT Pro Pro Ile Asp Glu Val Ile Ser Thr Pro Gly Val Val Ala Arg Phe 120 125 130 | 439 |
| GTG GAG TTC CTC AAA CGA AAA GAG AAT TGT TCA CTG CAG TTT GAA TCA Val Glu Phe Lys Lys Arg Lys Glu Asn Cys Ser Leu Gln Phe Glu Ser 135 140 145 | 487 |
| GCT TGG GTA CTG ACA AAT ATT GCT TCA GGA AAT TCT CTT CAG ACC CGA Ala Trp Val Leu Thr Asn Ile Ala Ser Gly Asn Ser Leu Gln Thr Arg 150 155 160 | 535 |
| ATT GTG ATT CAG GCA AGA GCT GTG CCC ATC TTC ATA GAG TTG CTC AGC Ile Val Ile Gln Ala Arg Ala Val Pro Ile Phe Ile Glu Leu Leu Ser 165 170 175 | 583 |
| TCA GAG TTT GAA GAT GTC CAG GAA CAG GCA GTC TGG GCT CTT GGC AAC Ser Glu Phe Glu Asp Val Gln Glu Gln Ala Val Trp Ala Leu Gly Asn 180 185 190 195 | 631 |
| ATT GCT GGA GAT AGT ACC ATG TGC AGG GAC TAT GTC TTA GAC TGC AAT Ile Ala Gly Asp Ser Thr Met Cys Arg Asp Tyr Val Leu Asp Cys Asn 200 205 210 | 679 |
| ATC CTT CCC CCT CTT TTG CAG TTA TTT TCA AAG CAA AAC CGC CTG ACC Ile Leu Pro Pro Leu Leu Gln Leu Phe Ser Lys Gln Asn Arg Leu Thr 215 220 225 | 727 |
| ATG ACC CGG AAT GCA GTA TGG GCT TTG TCT AAT CTC TGT AGA GGG AAA Met Thr Arg Asn Ala Val Trp Ala Leu Ser Asn Leu Cys Arg Gly Lys 230 235 240 | 775 |
| AGT CCA CCT CCA GAA TTT GCA AAG GTT TCT CCA TGT CTG AAT GTG CTT Ser Pro Pro Pro Glu Phe Ala Lys Val Ser Pro Cys Leu Asn Val Leu 245 250 255 | 823 |
| TCC TGG TTG CTG TTT GTC AGT GAC ACT GAT GTA CTG GCT GAT GCC TGC Ser Trp Leu Leu Phe Val Ser Asp Thr Asp Val Leu Ala Asp Ala Cys 260 265 270 275 | 871 |
| TGG GCC CTC TCA TAT CTA TCA GAT GGA CCC AAT GAT AAA ATT CAA GCG Trp Ala Leu Ser Tyr Leu Ser Asp Gly Pro Asn Asp Lys Ile Gln Ala 280 285 290 | 919 |
| GTC ATC GAT GCG GGA GTA TGT AGG AGA CTT GTG GAA CTG CTG ATG CAT Val Ile Asp Ala Gly Val Cys Arg Arg Leu Val Glu Leu Leu Met His 295 300 305 | 967 |
| AAT GAT TAT AAA GTG GTT TCT CCT GCT TTG CGA GCT GTG GGA AAC ATT Asn Asp Tyr Lys Val Val Ser Pro Ala Leu Arg Ala Val Gly Asn Ile 310 315 320 | 1015 |
| GTC ACA GGG GAT GAT ATT CAG ACA CAG GTA ATT CTG AAT TGC TCA GCT Val Thr Gly Asp Asp Ile Gln Thr Gln Val Ile Leu Asn Cys Ser Ala 325 330 335 | 1063 |

CTG CAG AGT TTA TTG CAT TTG CTG AGT AGC CCA AAG GAA TCT ATC AAA 1111
 Leu Gln Ser Leu Leu His Leu Leu Ser Ser Pro Lys Glu Ser Ile Lys 355
 340 345 350

AAG GAA GCA TGT TGG ACG ATA TCT AAT ATT ACA GCT GGA AAT AGG GCA 1159
 Lys Glu Ala Cys Trp Thr Ile Ser Asn Ile Thr Ala Gly Asn Arg Ala 370
 360 365

CAG ATC CAG ACT GTG ATA GAT GCC AAC ATT TTC CCA GCC CTC ATT AGT 1207
 Gln Ile Gln Thr Val Ile Asp Ala Asn Ile Phe Pro Ala Leu Ile Ser 385
 375 380

ATT TTA CAA ACT GCT GAA TTT CGG ACA AGA AAA GAA GCA GCT TGG GCC 1255
 Ile Leu Gln Thr Ala Glu Phe Arg Thr Arg Lys Glu Ala Ala Trp Ala 400
 390 395

ATC ACA AAT GCA ACT TCT GGA GGA TCA GCT GAA CAG ATC AAG TAC CTA 1303
 Ile Thr Asn Ala Thr Ser Gly Gly Ser Ala Glu Gln Ile Lys Tyr Leu 415
 405 410

GTA GAA CTG GGT TGT ATC AAG CCG CTC TGT GAT CTC CTC ACG GTC ATG 1351
 Val Glu Leu Gly Cys Ile Lys Pro Leu Cys Asp Leu Leu Thr Val Met 435
 420 425 430

GAC TCT AAG ATT GTA CAG GTT GCC CTA AAT GGC TTG GAA AAT ATC CTG 1399
 Asp Ser Lys Ile Val Gln Val Ala Leu Asn Gly Leu Glu Asn Ile Leu 450
 440 445

AGG CTT GGA GAA CAG GAA GCC AAA AGG AAC GGC ACT GGC ATT AAC CCT 1447
 Arg Leu Gly Glu Gln Glu Ala Lys Arg Asn Gly Thr Gly Ile Asn Pro 465
 455 460

TAC TGT GCT TTG ATT GAA GAA GCT TAT GGT CTG GAT AAA ATT GAG TTC 1495
 Tyr Cys Ala Leu Ile Glu Glu Ala Tyr Gly Leu Asp Lys Ile Glu Phe 480
 470 475

TTA CAG AGT CAT GAA AAC CAG GAG ATC TAC CAA AAG GCC TTT GAT CTT 1543
 Leu Gln Ser His Glu Asn Gln Glu Ile Tyr Gln Lys Ala Phe Asp Leu 495
 485 490

ATT GAG CAT TAC TTC GGG ACC GAA GAT GAA GAC AGC AGC ATT GCA CCC 1591
 Ile Glu His Tyr Phe Gly Thr Glu Asp Glu Asp Ser Ser Ile Ala Pro 515
 500 505 510

CAG GTT GAC CTT AAC CAG CAG CAG TAC ATC TTC CAA CAG TGT GAG GCT 1639
 Gln Val Asp Leu Asn Gln Gln Gln Tyr Ile Phe Gln Gln Cys Glu Ala 530
 520 525

CCT ATG GAA GGT TTC CAG CTT TGA AGCAATACTC TGCTTTCACG TACCTGTGCT 1693
 Pro Met Glu Gly Phe Gln Leu * 535

CAGACCAGGC TACCCAGTCG AGTCCTCTTG TGGAGCCCAC AGTCCTCATG GAGCTAACTT 1753

CTCAAATGTT TTCCATAATA CTGTTTGCGC TCATTTGCTT GCCTTGCGCA CCTGCTCTCT 1813

TACACACATC TGGAAAACCT CCGGCTCTCT GTGGTGGGAT ACCCTTCTAA TAAAAGGGTA 1873

ACCAGAACGG CCCACTCTCT TTTACGGAAA AATCCCTAGG CTTTGGAGAT CCGCACTTAC 1933

ATTAGAGTTA TGGGAATATA CACATATTAA TGTGGCTCCC TTTTCTTGT GGGGGAATAA 1993

AAGAGGACTC CTCCTCATT CTTTAACAT GGGGGAAAAA ACTGACATTA AAAGATGAGA 2053

CTAAATCTTT ATCTTGAATT TTACACAAC ACTTACGACA AGGGAGATGT TTAGACCTGT 2113

TGGTATACTT CAGAGTACTT TTCATGAGTT CTTCACAGT GAACCCTTGG ATTACCTGGT 2173

GGCTTTTCT AGCCAGATTG CATTAACTCT TACTGAGATT GGATGGTTTT CTTTCCTCTA 2233
TTGGCGCCAT TCTTCAGATA TTAAAGTTAA ACCATCCACT CCCTCACCTT CAGCCTTCAG 2293
TGAATGTGCT TTCTAGTTGT CAGGAATGCT GAAGAATTAA CACTTTGACT CCTAAATGTG 2353
ATACTGGTGG GTAAGAGCAG GGCACATTTA ATTTGTTCGC TTTTGCTTCT CTTTGGTCTG 2413
GGCACATTTA ATTTGTTCGC TTTTGCTTCT CTTTCGTCTT TTCGAATACT TAGTAATCGA 2473
AAACCATATC CTGTAATTTA ATAAAAA CTAAGGACGA AAAAAACCCCT CCAATTTTCC 2533
CAAATGCAAT CAGTGTAACCT AGGGGCTGTG TTTCTGCATT AAAATAAATG TTTCAGGCTT 2593
TGTGGTCCTG ATCAAGGTCC TCATTAAAAA ATTGGAGTTC ACCCTAGGCT TTTCCCCTCT 2653
GTGACTGGCA GATAACACAT ACTTTTGAAA GTAACCTTGG GATTTTTTTT CTTAGGTGCA 2713
GCTCGATTCT AATCTTTTCA TGCTGCACAC GATTCCTTTA ATCGATAGCA TCCTTATCTG 2773
AAAGAAATAA CCATCTTCTC AACATGACCT GCTTAACCCA AATAAGAACA GTGATCTTAT 2833
AACCTCATTG TTTCTAATC TATTTTATTT CATCTCTGCG TAGTACTGTG CCGCTTCCCC 2893
CTCCCCCAC ACAAAATAAA AACAGTATCT CGCTTCTGGC TCATTTT 2940

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 539 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met Thr Thr Pro Gly Lys Glu Asn Phe Arg Leu Lys Ser Tyr Lys Asn
1 5 10 15
Lys Ser Leu Asn Pro Asp Glu Met Arg Arg Arg Arg Glu Glu Glu Gly
20 25 30
Leu Gln Leu Arg Lys Gln Lys Arg Glu Glu Gln Leu Phe Lys Arg Arg
35 40 45
Asn Val Ala Thr Ala Glu Glu Glu Thr Glu Glu Glu Val Met Ser Asp
50 55 60
Gly Gly Phe His Glu Ala Gln Ile Ser Asn Met Glu Met Ala Pro Gly
65 70 75 80
Gly Val Ile Thr Ser Asp Met Ile Glu Met Ile Phe Ser Lys Ser Pro
85 90 95
Glu Gln Gln Leu Ser Ala Thr Gln Lys Phe Arg Lys Leu Leu Ser Lys
100 105 110
Glu Pro Asn Pro Pro Ile Asp Glu Val Ile Ser Thr Pro Gly Val Val
115 120 125
Ala Arg Phe Val Glu Phe Leu Lys Arg Lys Glu Asn Cys Ser Leu Gln
130 135 140
Phe Glu Ser Ala Trp Val Leu Thr Asn Ile Ala Ser Gly Asn Ser Leu
145 150 155 160

Gln Thr Arg Ile Val Ile Gln Ala Arg Ala Val Pro Ile Phe Ile Glu
 165 170 175
 Leu Leu Ser Ser Glu Phe Glu Asp Val Gln Glu Gln Ala Val Trp Ala
 180 185 190
 Leu Gly Asn Ile Ala Gly Asp Ser Thr Met Cys Arg Asp Tyr Val Leu
 195 200 205
 Asp Cys Asn Ile Leu Pro Pro Leu Leu Gln Leu Phe Ser Lys Gln Asn
 210 215 220
 Arg Leu Thr Met Thr Arg Asn Ala Val Trp Ala Leu Ser Asn Leu Cys
 225 230 235 240
 Arg Gly Lys Ser Pro Pro Pro Glu Phe Ala Lys Val Ser Pro Cys Leu
 245 250 255
 Asn Val Leu Ser Trp Leu Leu Phe Val Ser Asp Thr Asp Val Leu Ala
 260 265 270
 Asp Ala Cys Trp Ala Leu Ser Tyr Leu Ser Asp Gly Pro Asn Asp Lys
 275 280 285
 Ile Gln Ala Val Ile Asp Ala Gly Val Cys Arg Arg Leu Val Glu Leu
 290 295 300
 Leu Met His Asn Asp Tyr Lys Val Val Ser Pro Ala Leu Arg Ala Val
 305 310 315 320
 Gly Asn Ile Val Thr Gly Asp Asp Ile Gln Thr Gln Val Ile Leu Asn
 325 330 335
 Cys Ser Ala Leu Gln Ser Leu Leu His Leu Leu Ser Ser Pro Lys Glu
 340 345 350
 Ser Ile Lys Lys Glu Ala Cys Trp Thr Ile Ser Asn Ile Thr Ala Gly
 355 360 365
 Asn Arg Ala Gln Ile Gln Thr Val Ile Asp Ala Asn Ile Phe Pro Ala
 370 375 380
 Leu Ile Ser Ile Leu Gln Thr Ala Glu Phe Arg Thr Arg Lys Glu Ala
 385 390 395 400
 Ala Trp Ala Ile Thr Asn Ala Thr Ser Gly Gly Ser Ala Glu Gln Ile
 405 410 415
 Lys Tyr Leu Val Glu Leu Gly Cys Ile Lys Pro Leu Cys Asp Leu Leu
 420 425 430
 Thr Val Met Asp Ser Lys Ile Val Gln Val Ala Leu Asn Gly Leu Glu
 435 440 445
 Asn Ile Leu Arg Leu Gly Glu Gln Glu Ala Lys Arg Asn Gly Thr Gly
 450 455 460
 Ile Asn Pro Tyr Cys Ala Leu Ile Glu Glu Ala Tyr Gly Leu Asp Lys
 465 470 475 480
 Ile Glu Phe Leu Gln Ser His Glu Asn Gln Glu Ile Tyr Gln Lys Ala
 485 490 495
 Phe Asp Leu Ile Glu His Tyr Phe Gly Thr Glu Asp Glu Asp Ser Ser
 500 505 510
 Ile Ala Pro Gln Val Asp Leu Asn Gln Gln Gln Tyr Ile Phe Gln Gln

515

520

525

Cys Glu Ala Pro Met Glu Gly Phe Gln Leu *
 530 535

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 542 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS:
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Met Asp Asn Gly Thr Asp Ser Ser Thr Ser Lys Phe Val Pro Glu Tyr
 1 5 10 15
 Arg Arg Thr Asn Phe Lys Asn Lys Gly Arg Phe Ser Ala Asp Glu Leu
 20 25 30
 Arg Arg Arg Arg Asp Thr Gln Gln Val Glu Leu Arg Lys Ala Lys Arg
 35 40 45
 Asp Glu Ala Leu Ala Lys Arg Arg Asn Phe Ile Pro Pro Thr Asp Gly
 50 55 60
 Ala Asp Ser Asp Glu Glu Asp Glu Ser Ser Val Ser Ala Asp Gln Gln
 65 70 75 80
 Phe Tyr Ser Gln Leu Gln Gln Glu Leu Pro Gln Met Thr Gln Gln Leu
 85 90 95
 Asn Ser Asp Asp Met Gln Glu Gln Leu Ser Ala Thr Val Lys Phe Arg
 100 105 110
 Gln Ile Leu Ser Arg Glu His Arg Pro Pro Ile Asp Val Val Ile Gln
 115 120 125
 Ala Gly Val Val Pro Arg Leu Val Glu Phe Met Arg Glu Asn Gln Pro
 130 135 140
 Glu Met Leu Gln Leu Glu Ala Ala Trp Ala Leu Thr Asn Ile Ala Ser
 145 150 155 160
 Gly Thr Ser Ala Gln Thr Lys Val Val Val Asp Ala Asp Ala Val Pro
 165 170 175
 Leu Phe Ile Gln Leu Leu Tyr Thr Gly Ser Val Glu Val Lys Glu Gln
 180 185 190
 Ala Ile Trp Ala Leu Gly Asn Val Ala Gly Asp Ser Thr Asp Tyr Arg
 195 200 205
 Asp Tyr Val Leu Gln Cys Asn Ala Met Glu Pro Ile Leu Gly Leu Phe
 210 215 220
 Asn Ser Asn Lys Pro Ser Leu Ile Arg Thr Ala Thr Trp Thr Leu Ser
 225 230 235 240
 Asn Leu Cys Arg Gly Lys Lys Pro Gln Pro Asp Trp Ser Val Val Ser
 245 250 255

Gln Ala Leu Pro Thr Leu Ala Lys Leu Ile Tyr Ser Met Asp Thr Glu
 260 265 270
 Thr Leu Val Asp Ala Cys Trp Ala Ile Ser Tyr Leu Ser Asp Gly Pro
 275 280 285
 Gln Glu Ala Ile Gln Ala Val Ile Asp Val Arg Ile Pro Lys Arg Leu
 290 295 300
 Val Glu Leu Leu Ser His Glu Ser Thr Leu Val Gln Thr Pro Ala Leu
 305 310 315 320
 Arg Ala Val Gly Asn Ile Val Thr Gly Asn Asp Leu Gln Thr Gln Val
 325 330 335
 Val Ile Asn Ala Gly Val Leu Pro Ala Leu Arg Leu Leu Leu Ser Ser
 340 345 350
 Pro Lys Glu Asn Ile Lys Lys Glu Ala Cys Trp Thr Ile Ser Asn Ile
 355 360 365
 Thr Ala Gly Asn Thr Glu Gln Ile Gln Ala Val Ile Asp Ala Asn Leu
 370 375 380
 Ile Pro Pro Leu Val Lys Leu Leu Glu Val Ala Glu Tyr Lys Thr Lys
 385 390 395 400
 Lys Glu Ala Cys Trp Ala Ile Ser Asn Ala Ser Ser Gly Gly Leu Gln
 405 410 415
 Arg Pro Asp Ile Ile Arg Tyr Leu Val Ser Gln Gly Cys Ile Lys Pro
 420 425 430
 Leu Cys Asp Leu Leu Glu Ile Ala Asp Asn Arg Ile Ile Glu Val Thr
 435 440 445
 Leu Asp Ala Leu Glu Asn Ile Leu Lys Met Gly Glu Ala Asp Lys Glu
 450 455 460
 Ala Arg Gly Leu Asn Ile Asn Glu Asn Ala Asp Phe Ile Glu Lys Ala
 465 470 475 480
 Gly Gly Met Glu Lys Ile Phe Asn Cys Gln Gln Asn Glu Asn Asp Lys
 485 490 495
 Ile Tyr Glu Lys Ala Tyr Lys Ile Ile Glu Thr Tyr Phe Gly Glu Glu
 500 505 510
 Glu Asp Ala Val Asp Glu Thr Met Ala Pro Gln Asn Ala Gly Asn Thr
 515 520 525
 Phe Gly Phe Gly Ser Asn Val Asn Gln Gln Phe Asn Phe Asn
 530 535 540

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 170 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: unknown
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GGAGGCACCG AAGGGCAGCG CCGAGTCGGA GGGGGCGAAG ATTGACGCCA GTAAGAACGA 60
GGAGGATGAA GGCCATTCAA ACTCCTCCCC ACGACACTCT GAAGCAGCGA CGGCACAGCG 120
GGAAGAATGG AAAATGTTTA TAGGAGGCCT TAGCTGGGAC ACTACAAAGA 170

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1827 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: unknown
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA

- (ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1362

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GAG GTC AAT GTG GAG CTG AGG AAA GCT AAG AAG GAT GAC CAG ATG CTG 48
Glu Val Asn Val Glu Leu Arg Lys Ala Lys Lys Asp Asp Gln Met Leu
1 5 10 15
AAG AGG AGA AAT GTA AGC TCA TTT CCT GAT GAT GCT ACT TCT CCG CTG 96
Lys Arg Arg Asn Val Ser Ser Phe Pro Asp Asp Ala Thr Ser Pro Leu
20 25 30
CAG GAA AAC CGC AAC AAC CAG GGC ACT GTA AAT TGG TCT GTT GAT GAC 144
Gln Glu Asn Arg Asn Asn Gln Gly Thr Val Asn Trp Ser Val Asp Asp
35 40 45
ATT GTC AAA GGC ATA AAT AGC AGC AAT GTG GAA AAT CAG CTC CAA GCT 192
Ile Val Lys Gly Ile Asn Ser Ser Asn Val Glu Asn Gln Leu Gln Ala
50 55 60
ACT CAA GCT GCC AGG AAA CTA CTT TCC AGA GAA AAA CAG CCC CCC ATA 240
Thr Gln Ala Ala Arg Lys Leu Leu Ser Arg Glu Lys Gln Pro Pro Ile
65 70 75 80
GAC AAC ATA ATC CGG GCT GGT TTG ATT CCG AAA TTT GTG TCC TTC TTG 288
Asp Asn Ile Ile Arg Ala Gly Leu Ile Pro Lys Phe Val Ser Phe Leu
85 90 95
GGC AGA ACT GAT TGT AGT CCC ATT CAG TTT GAA TCT GCT TGG GCA CTC 336
Gly Arg Thr Asp Cys Ser Pro Ile Gln Phe Glu Ser Ala Trp Ala Leu
100 105 110
ACT AAC ATT GCT TCT GGG ACA TCA GAA CAA ACC AAG GCT GTG GTA GAT 384
Thr Asn Ile Ala Ser Gly Thr Ser Glu Gln Thr Lys Ala Val Val Asp
115 120 125
GGA GGT GCC ATC CCA GCA TTC ATT TCT CTG TTG GCA TCT CCC CAT GCT 432
Gly Gly Ala Ile Pro Ala Phe Ile Ser Leu Leu Ala Ser Pro His Ala
130 135 140
CAC ATC AGT GAA CAA GCT GTC TGG GCT CTA GGA AAC ATT GCA GGT GAT 480
His Ile Ser Glu Gln Ala Val Trp Ala Leu Gly Asn Ile Ala Gly Asp
145 150 155 160
GGC TCA GTG TTC CGA GAC TTG GTT ATT AAG TAC GGT GCA GTT GAC CCA 528

| | | | | |
|---|-----|-----|-----|------|
| Gly Ser Val Phe Arg Asp Leu Val Ile Lys Tyr Gly Ala Val Asp Pro | 165 | 170 | 175 | |
| CTG TTG GCT CTC CTT GCA GTT CCT GAT ATG TCA TCT TTA GCA TGT GGC | | | | 576 |
| Leu Leu Ala Leu Leu Ala Val Pro Asp Met Ser Ser Leu Ala Cys Gly | 180 | 185 | 190 | |
| TAC TTA CGT AAT CTT ACC TGG ACA CTT TCT AAT CTT TGC CGC AAC AAG | | | | 624 |
| Tyr Leu Arg Asn Leu Thr Trp Thr Leu Ser Asn Leu Cys Arg Asn Lys | 195 | 200 | 205 | |
| AAT CCT GCA CCC CCG ATA GAT GCT GTT GAG CAG ATT CTT CCT ACC TTA | | | | 672 |
| Asn Pro Ala Pro Pro Ile Asp Ala Val Glu Gln Ile Leu Pro Thr Leu | 210 | 215 | 220 | |
| GTT CGG CTC CTG CAT CAT GAT GAT CCA GAA GTG TTA GCA GAT ACC TGC | | | | 720 |
| Val Arg Leu Leu His His Asp Asp Pro Glu Val Leu Ala Asp Thr Cys | 225 | 230 | 235 | |
| TGG GCT ATT TCC TAC CTT ACT GAT GGT CCA AAT GAA CGA ATT GGC ATG | | | | 768 |
| Trp Ala Ile Ser Tyr Leu Thr Asp Gly Pro Asn Glu Arg Ile Gly Met | 245 | 250 | 255 | |
| GTG GTG AAA ACA GGA GTT GTG CCC CAA CTT GTG AAG CTT CTA GGA GCT | | | | 816 |
| Val Val Lys Thr Gly Val Val Pro Gln Leu Val Lys Leu Leu Gly Ala | 260 | 265 | 270 | |
| TCT GAA TTG CCA ATT GTG ACT CCT GCC CTA AGA GCC ATA GGG AAT ATT | | | | 864 |
| Ser Glu Leu Pro Ile Val Thr Pro Ala Leu Arg Ala Ile Gly Asn Ile | 275 | 280 | 285 | |
| GTC ACT GGT ACA GAT GAA CAG ACT CAG GTT GTG ATT GAT GCA GGA GCA | | | | 912 |
| Val Thr Gly Thr Asp Glu Gln Thr Gln Val Val Ile Asp Ala Gly Ala | 290 | 295 | 300 | |
| CTC GCC GTC TTT CCC AGC CTG CTC ACC AAC CCC AAA ACT AAC ATT CAG | | | | 960 |
| Leu Ala Val Phe Pro Ser Leu Leu Thr Asn Pro Lys Thr Asn Ile Gln | 305 | 310 | 315 | |
| AAG GAA GCT ACG TGG ACA ATG TCA AAC ATC ACA GCC GGC CGC CAG GAC | | | | 1008 |
| Lys Glu Ala Thr Trp Thr Met Ser Asn Ile Thr Ala Gly Arg Gln Asp | 325 | 330 | 335 | |
| CAG ATA CAG CAA GTT GTG AAT CAT GGA TTA GTC CCA TTC CTT GTC AGT | | | | 1056 |
| Gln Ile Gln Gln Val Val Asn His Gly Leu Val Pro Phe Leu Val Ser | 340 | 345 | 350 | |
| GTT CTC TCT AAG GCA GAT TTT AAG ACA CAA AAG GAA GCT GTG TGG GCC | | | | 1104 |
| Val Leu Ser Lys Ala Asp Phe Lys Thr Gln Lys Glu Ala Val Trp Ala | 355 | 360 | 365 | |
| GTG ACC AAC TAT ACC AGT GGT GGA ACA GTT GAA CAG ATT GTG TAC CTT | | | | 1152 |
| Val Thr Asn Tyr Thr Ser Gly Gly Thr Val Glu Gln Ile Val Tyr Leu | 370 | 375 | 380 | |
| GTT CAC TGT GGC ATA ATA GAA CCG TTG ATG AAC CTC TTA ACT GCA AAA | | | | 1200 |
| Val His Cys Gly Ile Ile Glu Pro Leu Met Asn Leu Leu Thr Ala Lys | 385 | 390 | 395 | |
| GAT ACC AAG ATT ATT CTG GTT ATC CTG GAT GCC ATT TCA AAT ATC TTT | | | | 1248 |
| Asp Thr Lys Ile Ile Leu Val Ile Leu Asp Ala Ile Ser Asn Ile Phe | 405 | 410 | 415 | |
| CAG GCT GCT GAG AAA CTA GGT GAA ACT AGC TGC CCG TCT TCA CAG ATT | | | | 1296 |
| Gln Ala Ala Glu Lys Leu Gly Glu Thr Ser Cys Pro Ser Ser Gln Ile | 420 | 425 | 430 | |

CAA GAA CAA GGG AAA AGA CAG TAC AGA AAT GAG GCG TCC GAG GCG TCG 1344
 Gln Glu Gln Gly Lys Arg Gln Tyr Arg Asn Glu Ala Ser Glu Ala Ser
 435 440 445

CAG AAT AGA GAA ACT TAG TATAATGATT GAAGAATGTG GAGGCTTAGA 1392
 Gln Asn Arg Glu Thr *
 450

CAAAATTGAA GCTCTACAAA ACCATGAAAA TGAGTCTGTG TATAAGGCTT CGTTAAGCTT 1452
 AATTGAGAAG TATTTCTCTG TAGAGGAAGA GGAAGATCAA AACGTTGTAC CAGAAACTAC 1512
 CTCTGAAGGC TACACTTTCC AAGTTCAGGA TGGGGCTCCT GGGACCTTTA ACTTTTAGAT 1572
 CATGTAGCTG AGACATAAAT TTGTTGTGTA CTACGTTTGG TATTTTGTCT TATTGTTTCT 1632
 CTACTAAGAA CTCTTTCTTA AATGTGGTTT GTTACTGTAG CACTTTTTTAC ACTGAAACTA 1692
 TACTTGAACA GTTCCAACG TACATACATA CTGTATGAAG CTTGTCCTCT GACTAGGTTT 1752
 CTAATTTCTA TGTGGAATTT CCTATCTTGC AGCATCCTGT AAATAAACAT TCAAGTCCAC 1812
 CCTTTTCTTG ACTTC 1827

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 454 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Glu Val Asn Val Glu Leu Arg Lys Ala Lys Lys Asp Asp Gln Met Leu
 1 5 10 15

Lys Arg Arg Asn Val Ser Ser Phe Pro Asp Asp Ala Thr Ser Pro Leu
 20 25 30

Gln Glu Asn Arg Asn Asn Gln Gly Thr Val Asn Trp Ser Val Asp Asp
 35 40 45

Ile Val Lys Gly Ile Asn Ser Ser Asn Val Glu Asn Gln Leu Gln Ala
 50 55 60

Thr Gln Ala Ala Arg Lys Leu Leu Ser Arg Glu Lys Gln Pro Pro Ile
 65 70 75 80

Asp Asn Ile Ile Arg Ala Gly Leu Ile Pro Lys Phe Val Ser Phe Leu
 85 90 95

Gly Arg Thr Asp Cys Ser Pro Ile Gln Phe Glu Ser Ala Trp Ala Leu
 100 105 110

Thr Asn Ile Ala Ser Gly Thr Ser Glu Gln Thr Lys Ala Val Val Asp
 115 120 125

Gly Gly Ala Ile Pro Ala Phe Ile Ser Leu Leu Ala Ser Pro His Ala
 130 135 140

His Ile Ser Glu Gln Ala Val Trp Ala Leu Gly Asn Ile Ala Gly Asp
 145 150 155 160

Gly Ser Val Phe Arg Asp Leu Val Ile Lys Tyr Gly Ala Val Asp Pro

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

| | |
|--|-----|
| GAACGACCAA GAGGGTGTTC GACTGCTAGA GCCGAGCAGA AGCGTGCCTA AATCAAAGGA | 60 |
| ACTTGTCTTCT TCAAGCTCTT CTGGCAGTGA TTCTGACAGT GAGGTTGACA AAAAGTTAAG | 120 |
| CAGGAAAAAG CAAGTTGCTC CAGAAAAACC TGTAAGAAA CAAAAGACAG GTGAGACTTC | 180 |
| GAGAGCCCTG TCATCTTCTA AACAGAGCAG CAGCAGCAGA GATGATAACA TGTTTCAGAT | 240 |
| TGGGAAAATG AGGTCAGTT | 259 |

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 221 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: unknown
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

| | |
|---|-----|
| TGTCGACTGT GGCTTTGAGC ATCCGTCAGA AGTCCAGCAT GAGTGCATCC CTCAGGCCAT | 60 |
| TCTGGGAATG GATGTCCTGT GCCAGGCCAA GTCGGGCATG GGAAAGACAG CAGTGTCTGT | 120 |
| CTTGCCACA CTGCAACAGC TGGAGCCAGT TACTGGGCAG GTGTCTGTAC TGGTGATGTG | 180 |
| TCACACTCGG GAGTTGGCTT TTCAGATCAG CAAGGAATAT G | 221 |

(2) INFORMATION FOR SEQ ID NO:18:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 372 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: unknown
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

| | |
|---|-----|
| ATTTGTAAAC CCCGGAGCGA GGTTCGTCTT ACCCGAGGCC GCTGCTGTGC GGAGACCCCC | 60 |
| GGGTGAAGCC ACCGTCATCA TGTCTGACCA GGAGGCAAAA CCTTCAACTG AGGACTTGGG | 120 |
| GGATAAGAAG GAAGGTGAAT ATATTAACT CAAAGTCATT GGACAGGATA GCAGTGAGAT | 180 |
| TCACTTCAAA GTGAAAATGA CAACACATCT CAAGAACTC AAAGAATCAT ACTGTCAAAG | 240 |
| ACAGGGTGTT CCAATGAATT CACTCAGGTT TCTCTTTGAG GGTGAGAGAA TTGCTGATAA | 300 |
| TCATACTCCA AAAGAACTGG GAATGGAGGA AGAAGTTGTG ATTGAAGTTT ATCAGGAACA | 360 |
| AACGGGGGGT CA | 372 |

(2) INFORMATION FOR SEQ ID NO:19:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 2675 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: unknown
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA

- (ix) FEATURE:
 (A) NAME/KEY: CDS
 (B) LOCATION: 104..2311

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

| | |
|---|-----|
| TCTGACCCTC GTCCCGCCCC CGCCATTCGC CGCCTCCTCC TGTCCCGCAG TCGGCGTCCA | 60 |
| CGGGCTCTGC TTGTTTCGTGT GTGTGTCGTT GCAGGCCTTA TTC ATG GGC TCA CCG | 115 |
| Met Gly Ser Pro | |
| 1 | |
| CTG AGG TTC GAC GGG CGG GTG GTA CTG GTC ACC GGC GCG GGG GCA GGA | 163 |
| Leu Arg Phe Asp Gly Arg Val Val Leu Val Thr Gly Ala Gly Ala Gly | |
| 5 10 15 20 | |
| TTG GGC CGA GCC TAT GCC CTG GCT TTT GCA GAA AGA GGA GCG TTA GTT | 211 |
| Leu Gly Arg Ala Tyr Ala Leu Ala Phe Ala Glu Arg Gly Ala Leu Val | |
| 25 30 35 | |
| GTT GTG AAT GAT TTG GGA GGG GAC TTC AAA GGA GTT GGT AAA GGC TCC | 259 |
| Val Val Asn Asp Leu Gly Gly Asp Phe Lys Gly Val Gly Lys Gly Ser | |
| 40 45 50 | |
| TTA GCT GAT AAG GTT GTT GAA GAA ATA AGA AGG AGA GGT GGA AAA GCA | 307 |
| Leu Ala Asp Lys Val Val Glu Glu Ile Arg Arg Arg Gly Gly Lys Ala | |
| 55 60 65 | |
| GTG GCC AAC TAT GAT TCA GTG GAA GAA GGA GAG AAG GTT GTG AAG ACA | 355 |
| Val Ala Asn Tyr Asp Ser Val Glu Glu Gly Glu Lys Val Val Lys Thr | |
| 70 75 80 | |
| GCC CTG GAT GCT TTT GGA AGA ATA GAT GTT GTG GTC AAC AAT GCT GGA | 403 |
| Ala Leu Asp Ala Phe Gly Arg Ile Asp Val Val Asn Asn Ala Gly | |
| 85 90 95 100 | |
| ATT CTG AGG GAT CAT TCC TTT GCT AGG ATA AGT GAT GAA GAC TGG GAT | 451 |
| Ile Leu Arg Asp His Ser Phe Ala Arg Ile Ser Asp Glu Asp Trp Asp | |
| 105 110 115 | |
| ATA ATC CAC AGA GTT CAT TTG CGG GGT TCA TTC CAA GTG ACA CGG GCA | 499 |
| Ile Ile His Arg Val His Leu Arg Gly Ser Phe Gln Val Thr Arg Ala | |
| 120 125 130 | |
| GCA TGG GAA CAC ATG AAG AAA CAG AAG TAT GGA AGG ATT ATT ATG ACT | 547 |
| Ala Trp Glu His Met Lys Lys Gln Lys Tyr Gly Arg Ile Ile Met Thr | |
| 135 140 145 | |
| TCA TCA GCT TCA GGA ATA TAT GGC AAC TTT GGC CAG GCC AAT TAT AGT | 595 |
| Ser Ser Ala Ser Gly Ile Tyr Gly Asn Phe Gly Gln Ala Asn Tyr Ser | |
| 150 155 160 | |
| GCT GCA AAG TTG GGT CTT CTG GGC CTT GCA AAT TCT CTT GCA ATT GAA | 643 |
| Ala Ala Lys Leu Gly Leu Leu Gly Leu Ala Asn Ser Leu Ala Ile Glu | |
| 165 170 175 180 | |
| GGC AGG AAA AGC AAC ATT CAT TGT AAC ACC ATT GCT CCT AAT GCC GGA | 691 |

| | | | | | | | | | | | | | | | | | |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|------|
| Gly | Arg | Lys | Ser | Asn | Ile | His | Cys | Asn | Thr | Ile | Ala | Pro | Asn | Ala | Gly | | |
| | | | | 185 | | | | | 190 | | | | | 195 | | | |
| TCA | CGG | ATG | ACT | CAG | ACA | GTT | ATG | CCT | GAA | GAT | CTT | GTG | GAA | GCC | TTG | | 739 |
| Ser | Arg | Met | Thr | Gln | Thr | Val | Met | Pro | Glu | Asp | Leu | Val | Glu | Ala | Leu | | |
| | | | 200 | | | | | 205 | | | | | 210 | | | | |
| AAG | CCA | GAG | TAT | GTG | GCA | CCT | CTT | GTC | CTT | TGG | CTT | TGT | CAC | GAG | AGT | | 787 |
| Lys | Pro | Glu | Tyr | Val | Ala | Pro | Leu | Val | Leu | Trp | Leu | Cys | His | Glu | Ser | | |
| | | 215 | | | | | 220 | | | | | 225 | | | | | |
| TGT | GAG | GAG | AAT | GGT | GGC | TTG | TTT | GAG | GTT | GGT | GCA | GGA | TGG | ATT | GGA | | 835 |
| Cys | Glu | Glu | Asn | Gly | Gly | Leu | Phe | Glu | Val | Gly | Ala | Gly | Trp | Ile | Gly | | |
| | 230 | | | | | 235 | | | | | 240 | | | | | | |
| AAA | TTA | CGC | TGG | GAG | CGG | ACT | CTT | GGA | GCT | ATT | GTA | AGA | CAA | AAG | AAT | | 883 |
| Lys | Leu | Arg | Trp | Glu | Arg | Thr | Leu | Gly | Ala | Ile | Val | Arg | Gln | Lys | Asn | | |
| | 245 | | | | 250 | | | | 255 | | | | | | 260 | | |
| CAC | CCA | ATG | ACT | CCT | GAG | GCA | GTC | AAG | GCT | AAC | TGG | AAG | AAG | ATC | TGT | | 931 |
| His | Pro | Met | Thr | Pro | Glu | Ala | Val | Lys | Ala | Asn | Trp | Lys | Lys | Ile | Cys | | |
| | | | | 265 | | | | | 270 | | | | | 275 | | | |
| GAC | TTT | GAG | AAT | GCC | AGC | AAG | CCT | CAG | AGT | ATC | CAA | GAA | TCA | ACT | GGC | | 979 |
| Asp | Phe | Glu | Asn | Ala | Ser | Lys | Pro | Gln | Ser | Ile | Gln | Glu | Ser | Thr | Gly | | |
| | | | 280 | | | | | 285 | | | | | 290 | | | | |
| AGT | ATA | ATT | GAA | GTT | CTG | AGT | AAA | ATA | GAT | TCA | GAA | GGA | GGA | GTT | TCA | | 1027 |
| Ser | Ile | Ile | Glu | Val | Leu | Ser | Lys | Ile | Asp | Ser | Glu | Gly | Gly | Val | Ser | | |
| | | 295 | | | | | 300 | | | | | 305 | | | | | |
| GCA | AAT | CAT | ACT | AGT | CGT | GCA | ACG | TCT | ACA | GCA | ACA | TCA | GGA | TTT | GCT | | 1075 |
| Ala | Asn | His | Thr | Ser | Arg | Ala | Thr | Ser | Thr | Ala | Thr | Ser | Gly | Phe | Ala | | |
| | 310 | | | | | 315 | | | | | 320 | | | | | | |
| GGA | GCT | ATT | GGC | CAG | AAA | CTC | CCT | CCA | TTT | TCT | TAT | GCT | TAT | ACG | GAA | | 1123 |
| Gly | Ala | Ile | Gly | Gln | Lys | Leu | Pro | Pro | Phe | Ser | Tyr | Ala | Tyr | Thr | Glu | | |
| | 325 | | | | 330 | | | | | 335 | | | | | 340 | | |
| CTG | GAA | GCT | ATT | ATG | TAT | GCC | CTT | GGA | GTG | GGA | GCG | TCA | ATC | AAG | GAT | | 1171 |
| Leu | Glu | Ala | Ile | Met | Tyr | Ala | Leu | Gly | Val | Gly | Ala | Ser | Ile | Lys | Asp | | |
| | | | | 345 | | | | | 350 | | | | | 355 | | | |
| CCA | AAA | GAT | TTG | AAA | TTT | ATT | TAT | GAA | GGA | AGT | TCT | GAT | TTC | TCC | TGT | | 1219 |
| Pro | Lys | Asp | Leu | Lys | Phe | Ile | Tyr | Glu | Gly | Ser | Ser | Asp | Phe | Ser | Cys | | |
| | | | 360 | | | | | 365 | | | | | 370 | | | | |
| TTG | CCC | ACC | TTC | GGA | GTT | ATC | ATA | GGT | CAG | AAA | TCT | ATG | ATG | GGT | GGA | | 1267 |
| Leu | Pro | Thr | Phe | Gly | Val | Ile | Ile | Gly | Gln | Lys | Ser | Met | Met | Gly | Gly | | |
| | | 375 | | | | | 380 | | | | | 385 | | | | | |
| GGA | TTA | GCA | GAA | ATT | CCT | GGA | CTT | TCA | ATC | AAC | TTT | GCA | AAG | GTT | CTT | | 1315 |
| Gly | Leu | Ala | Glu | Ile | Pro | Gly | Leu | Ser | Ile | Asn | Phe | Ala | Lys | Val | Leu | | |
| | 390 | | | | | 395 | | | | | 400 | | | | | | |
| CAT | GGA | GAG | CAG | TAC | TTA | GAG | TTA | TAT | AAA | CCA | CTT | CCC | AGA | GCA | GGA | | 1363 |
| His | Gly | Glu | Gln | Tyr | Leu | Glu | Leu | Tyr | Lys | Pro | Leu | Pro | Arg | Ala | Gly | | |
| | 405 | | | | 410 | | | | | 415 | | | | | 420 | | |
| AAA | TTA | AAA | TGT | GAA | GCA | GTT | GTT | GCT | GAT | GTC | CTA | GAT | AAA | GGA | TCC | | 1411 |
| Lys | Leu | Lys | Cys | Glu | Ala | Val | Val | Ala | Asp | Val | Leu | Asp | Lys | Gly | Ser | | |
| | | | | 425 | | | | 430 | | | | | | 435 | | | |
| GGT | GTA | GTG | ATT | ATT | ATG | GAT | GTC | TAT | TCT | TAT | TCT | GAG | AAG | GAA | CTT | | 1459 |
| Gly | Val | Val | Ile | Ile | Met | Asp | Val | Tyr | Ser | Tyr | Ser | Glu | Lys | Glu | Leu | | |
| | | | 440 | | | | | 445 | | | | | 450 | | | | |

| | |
|---|------|
| ATA TGC CAC AAT CAG TTC TCT CTC TTT CTT GTT GGC TCT GGA GGC TTT Ile Cys His Asn Gln Phe Ser Leu Phe Leu Val Gly Ser Gly Gly Phe 455 460 465 | 1507 |
| GGT GGA AAA CGG ACA TCA GAC AAA GTC AAG GTA GCT GTA GCC ATA CCT Gly Gly Lys Arg Thr Ser Asp Lys Val Lys Val Ala Val Ala Ile Pro 470 475 480 | 1555 |
| AAT AGA CCT CCT GAT GCT GTA CTT ACA GAT ACC ACC TCT CTT AAT CAG Asn Arg Pro Pro Asp Ala Val Leu Thr Asp Thr Thr Ser Leu Asn Gln 485 490 495 500 | 1603 |
| GCT GCT TTG TAC CGC CTC AGT GGA GAC CGG AAT CCC TTA CAC ATT GAT Ala Ala Leu Tyr Arg Leu Ser Gly Asp Arg Asn Pro Leu His Ile Asp 505 510 515 | 1651 |
| CCT AAC TTT GCT AGT CTA GCA GGT TTT GAC AAG CCC ATA TTA CAT GGA Pro Asn Phe Ala Ser Leu Ala Gly Phe Asp Lys Pro Ile Leu His Gly 520 525 530 | 1699 |
| TTA TGT ACA TTT GGA TTT TCT GCC AGG CGT GTG TTA CAG CAG TTT GCA Leu Cys Thr Phe Gly Phe Ser Ala Arg Arg Val Leu Gln Gln Phe Ala 535 540 545 | 1747 |
| GAT AAT GAT GTG TCA AGA TTC AAG GCA GTT AAG GCT CGT TTT GCA AAA Asp Asn Asp Val Ser Arg Phe Lys Ala Val Lys Ala Arg Phe Ala Lys 550 555 560 | 1795 |
| CCA GTA TAT CCA GGA CAA ACT CTA CAA ACT GAG ATG TGG AAG GAA GGA Pro Val Tyr Pro Gly Gln Thr Leu Gln Thr Glu Met Trp Lys Glu Gly 565 570 575 580 | 1843 |
| AAC AGA ATT CAT TTT CAA ACC AAG GTC CAA GAA ACT GGA GAC ATT GTC Asn Arg Ile His Phe Gln Thr Lys Val Gln Glu Thr Gly Asp Ile Val 585 590 595 | 1891 |
| ATT TCA AAT GCA TAT GTG GAT CTT GCA CCA ACA TCT GGT ACT TCA GCT Ile Ser Asn Ala Tyr Val Asp Leu Ala Pro Thr Ser Gly Thr Ser Ala 600 605 610 | 1939 |
| AAG ACA CCC TCT GAG GGC GGG AAG CTT CAG AGT ACC TTT GTA TTT GAG Lys Thr Pro Ser Glu Gly Gly Lys Leu Gln Ser Thr Phe Val Phe Glu 615 620 625 | 1987 |
| GAA ATA GGA CGC CGC CTA AAG GAT ATT GGG CCT GAG GTG GTG AAG AAA Glu Ile Gly Arg Arg Leu Lys Asp Ile Gly Pro Glu Val Val Lys Lys 630 635 640 | 2035 |
| GTA AAT GCT GTA TTT GAG TGG CAT ATA ACC AAA GGC GGA AAT ATT GGG Val Asn Ala Val Phe Glu Trp His Ile Thr Lys Gly Gly Asn Ile Gly 645 650 655 660 | 2083 |
| GCT AAG TGG ACT ATT GAC CTG AAA AGT GGT TCT GGA AAA GTG TAC CAA Ala Lys Trp Thr Ile Asp Leu Lys Ser Gly Ser Gly Lys Val Tyr Gln 665 670 675 | 2131 |
| GGC CCT GCA AAA GGT GCT GCT GAT ACA ACA ATC ATA CTT TCA GAT GAA Gly Pro Ala Lys Gly Ala Ala Asp Thr Thr Ile Ile Leu Ser Asp Glu 680 685 690 | 2179 |
| GAT TTC ATG GAG GTG GTC CTG GGC AAG CTT GAC CCT CAG AAG GCA TTC Asp Phe Met Glu Val Val Leu Gly Lys Leu Asp Pro Gln Lys Ala Phe 695 700 705 | 2227 |
| TTT AGT GGC AGG CTG AAG GCC AGA GGG AAC ATC ATG CTG AGC CAG AAA Phe Ser Gly Arg Leu Lys Ala Arg Gly Asn Ile Met Leu Ser Gln Lys 710 715 720 | 2275 |

CTT CAG ATG ATT CTT AAA GAC TAC GCC AAG CTC TGA AGGGCACACT 2321
 Leu Gln Met Ile Leu Lys Asp Tyr Ala Lys Leu *
 725 730 735

ACACTATTAA TAAAAATGGA ATCATTAAAT ACTCTCTTCA CCCAAATATG CTTGATTATT 2381
 CTGCAAAAGT GATTAGAACT AAGATGCAGG GGAAATTGCT TAACATTTTC AGATATCAGA 2441
 TAACTGCAGA TTTTCATTTT CTAATAATTT TTCATGTATC ATTATTTTTC CAAGGAACTA 2501
 TATATAAGCT AGCACATAAT TATCCTTCTG TTCTTAGATC TGTATCTTCA TAATAAAAAA 2561
 ATTTTGCCCA AGTCCTGTTT CCTTAGAATT TGTGATAGCA TTGATAAGTT GAAAGGAAAA 2621
 TTAAATCAAT AAAGGCCTTT GATACCTTTA AAAAAAAAAA AAAAAAAAAA AAAA 2675

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 736 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Met Gly Ser Pro Leu Arg Phe Asp Gly Arg Val Val Leu Val Thr Gly
 1 5 10 15

Ala Gly Ala Gly Leu Gly Arg Ala Tyr Ala Leu Ala Phe Ala Glu Arg
 20 25 30

Gly Ala Leu Val Val Val Asn Asp Leu Gly Gly Asp Phe Lys Gly Val
 35 40 45

Gly Lys Gly Ser Leu Ala Asp Lys Val Val Glu Glu Ile Arg Arg Arg
 50 55 60

Gly Gly Lys Ala Val Ala Asn Tyr Asp Ser Val Glu Glu Gly Glu Lys
 65 70 75 80

Val Val Lys Thr Ala Leu Asp Ala Phe Gly Arg Ile Asp Val Val Val
 85 90 95

Asn Asn Ala Gly Ile Leu Arg Asp His Ser Phe Ala Arg Ile Ser Asp
 100 105 110

Glu Asp Trp Asp Ile Ile His Arg Val His Leu Arg Gly Ser Phe Gln
 115 120 125

Val Thr Arg Ala Ala Trp Glu His Met Lys Lys Gln Lys Tyr Gly Arg
 130 135 140

Ile Ile Met Thr Ser Ser Ala Ser Gly Ile Tyr Gly Asn Phe Gly Gln
 145 150 155 160

Ala Asn Tyr Ser Ala Ala Lys Leu Gly Leu Leu Gly Leu Ala Asn Ser
 165 170 175

Leu Ala Ile Glu Gly Arg Lys Ser Asn Ile His Cys Asn Thr Ile Ala
 180 185 190

Pro Asn Ala Gly Ser Arg Met Thr Gln Thr Val Met Pro Glu Asp Leu
 195 200 205

Val Glu Ala Leu Lys Pro Glu Tyr Val Ala Pro Leu Val Leu Trp Leu
 210 215 220
 Cys His Glu Ser Cys Glu Glu Asn Gly Gly Leu Phe Glu Val Gly Ala
 225 230 235 240
 Gly Trp Ile Gly Lys Leu Arg Trp Glu Arg Thr Leu Gly Ala Ile Val
 245 250 255
 Arg Gln Lys Asn His Pro Met Thr Pro Glu Ala Val Lys Ala Asn Trp
 260 265 270
 Lys Lys Ile Cys Asp Phe Glu Asn Ala Ser Lys Pro Gln Ser Ile Gln
 275 280 285
 Glu Ser Thr Gly Ser Ile Ile Glu Val Leu Ser Lys Ile Asp Ser Glu
 290 295 300
 Gly Gly Val Ser Ala Asn His Thr Ser Arg Ala Thr Ser Thr Ala Thr
 305 310 315 320
 Ser Gly Phe Ala Gly Ala Ile Gly Gln Lys Leu Pro Pro Phe Ser Tyr
 325 330 335
 Ala Tyr Thr Glu Leu Glu Ala Ile Met Tyr Ala Leu Gly Val Gly Ala
 340 345 350
 Ser Ile Lys Asp Pro Lys Asp Leu Lys Phe Ile Tyr Glu Gly Ser Ser
 355 360 365
 Asp Phe Ser Cys Leu Pro Thr Phe Gly Val Ile Ile Gly Gln Lys Ser
 370 375 380
 Met Met Gly Gly Gly Leu Ala Glu Ile Pro Gly Leu Ser Ile Asn Phe
 385 390 395 400
 Ala Lys Val Leu His Gly Glu Gln Tyr Leu Glu Leu Tyr Lys Pro Leu
 405 410 415
 Pro Arg Ala Gly Lys Leu Lys Cys Glu Ala Val Val Ala Asp Val Leu
 420 425 430
 Asp Lys Gly Ser Gly Val Val Ile Ile Met Asp Val Tyr Ser Tyr Ser
 435 440 445
 Glu Lys Glu Leu Ile Cys His Asn Gln Phe Ser Leu Phe Leu Val Gly
 450 455 460
 Ser Gly Gly Phe Gly Gly Lys Arg Thr Ser Asp Lys Val Lys Val Ala
 465 470 475 480
 Val Ala Ile Pro Asn Arg Pro Pro Asp Ala Val Leu Thr Asp Thr Thr
 485 490 495
 Ser Leu Asn Gln Ala Ala Leu Tyr Arg Leu Ser Gly Asp Arg Asn Pro
 500 505 510
 Leu His Ile Asp Pro Asn Phe Ala Ser Leu Ala Gly Phe Asp Lys Pro
 515 520 525
 Ile Leu His Gly Leu Cys Thr Phe Gly Phe Ser Ala Arg Arg Val Leu
 530 535 540
 Gln Gln Phe Ala Asp Asn Asp Val Ser Arg Phe Lys Ala Val Lys Ala
 545 550 555 560
 Arg Phe Ala Lys Pro Val Tyr Pro Gly Gln Thr Leu Gln Thr Glu Met

